



TITLE:

Effect of active exploration of 3-D object views on the view-matching process in object recognition

AUTHOR(S):

Sasaoka, Takafumi; Asakura, Nobuhiko; Kawahara, Tetsuo

CITATION:

Sasaoka, Takafumi ...[et al]. Effect of active exploration of 3-D object views on the view-matching process in object recognition. Perception 2010, 39(3): 289-308

ISSUE DATE:

2010-03

URL:

<http://hdl.handle.net/2433/139452>

RIGHT:

© 2011 a Pion publication; This is not the published version. Please cite only the published version.; この論文は出版社版ではありません。引用の際には出版社版をご確認ご利用ください。

Effect of active exploration of 3-D object views on the view matching process in object recognition

Takafumi Sasaoka*

Department of Intelligence Science and Technology, Graduate School of Informatics, Kyoto
University

Human Information System Laboratory, Kanazawa Institute of Technology

Nobuhiko Asakura

Tetsuo Kawahara

Human Information System Laboratory, Kanazawa Institute of Technology

*Corresponding author

Department of Intelligence Science and Technology,
Graduate School of Informatics, Kyoto University
Yoshida-Hommachi
Sakyo-ku, Kyoto, 606-8501
Japan
Telephone & Fax: +81-75-753-3148
e-mail: sasaoka@i.kyoto-u.ac.jp

Effect of active exploration of 3-D object views on the view matching process in object recognition

Abstract

Harman et al (1999) showed that active exploration of novel objects allows faster recognition of these objects compared to passive viewing. In the present study, we examined whether active exploration of particular objects can facilitate subsequent view matching of different objects within the same category. In the generalization phase, participants performed a matching task between two serially presented views of novel paper-clip objects. When the two views were of the same object, these were related by rotation about the vertical or horizontal axis. Subsequently, in the observation phase, participants were presented with another five paper-clip objects. One group of participants (active group) actively explored each of the objects for 20 sec using a trackball over a limited range (-45° to 45°) of the horizontal axis. The other participants observed a replay of the active exploration of each of these objects by a participant in the active group (passive group) or observed a static view of the objects (still group). Following this observation phase, the generalization phase was repeated twice for all groups of participants. We found that the active group showed a clear expansion of the generalization range about the horizontal axis. However, this expansion disappeared when we replicated the experiment using a smaller range of active exploration (-30° to 30°) and the experiment in which the participants observed a

rotating object within the range of -45° to 45° . Furthermore, we found a significant improvement in view generalization about the vertical axis when participants actively explored objects using a turntable over a range in -45° to 45° of the vertical axis. Most importantly, since participants explored different objects from those used in the generalization phase, these results cannot be explained by improvement of the encoding process. Accordingly, the present findings suggest the following: 1) active exploration facilitates the process of view matching, 2) this effect depends on the exploration range and rotational axis, and 3) interaction between motor and visual information is necessary to facilitate the process of view matching.

1. Introduction

Human 3-D object recognition is robust with respect to an infinite variety of object appearances depending upon the viewpoint of the observer (object constancy). One of the main issues in the literature on object recognition research is clarification of the recognition system enabling object constancy. In daily life, we observe changes in the appearances of 3-D objects due to the manipulation of objects (e.g. picking up a tea cup and lifting it to the mouth) or moving ourselves to other positions relative to objects. These experiences might provide opportunities to acquire knowledge concerning view-transformation rules regarding 3-D objects leading to object constancy.

A number of researchers have pointed out the existence of interaction between visual information processing and active movement. Held and Hein (1963) showed that visual experiences, coordinated with active movement, are critical for development of visuo-spatial ability in neonatal kittens. Visuo-motor coordination is also important for human development (Piaget and Inhelder, 1948; Gibson, 1979). Wexler et al. (1998) demonstrated that mental rotation is facilitated when the directions of mental and motor rotation are compatible (see also Wohlschläger and Wohlschläger, 1998). Tong, Marlin, and Frost (1995) and Christou and Bühlhoff (1999) found that active learning of a visual scene can facilitate its later recognition.

These findings suggest the possibility that object recognition can be facilitated by information from active movement. The first study to address this possibility directly was that of Harman et al (1999). In their experiment, one set of computer-generated

novel 3-D objects was learned by active exploration, in which participants could rotate an object three-dimensionally using a trackball (learning phase). The other set of objects was learned by passive observation of the replay of another participant's active exploration. After the learning phase, participants performed an old/new recognition task. Harman et al showed that reaction time was significantly shorter when participants recognized objects learned by active manual control compared to learning by passive observation. They further argued that "active exploration could allow subjects to test 'predictions' about the expected deformations in the image that would occur when the object is rotated in a particular way" (Harman et al 1999, pp. 1316).

We note that object recognition can be characterized as involving two components. One is an encoding process to acquire object representations in the brain, and the other is a matching process, by which an input view and object representations are compared. The work by Harman et al. (1999) may be interpreted as providing evidence that active exploration can facilitate the matching process rather than the encoding process. However, in their experiments, an explicit old/new task was adopted, that is, the same set of objects was used in both the learning and recognition phases, which might have allowed participants to memorize particular views during the learning phase. To assess this possibility, James et al. (2001) adopted a perceptual matching task in which participants were presented with two novel object views related by rotation and judged whether these views were of the same object or not. They showed that reaction time in matching a 90° rotation view for actively studied objects was faster than for

passively studied objects. This finding suggests that the matching process is facilitated by active exploration of object views. However, since the same set of objects was used in both the familiarization and test phases, the possibility remains that “rather than facilitating mental rotation from one view to another, the active exploration condition may have created stronger associations between some sort of object template and different views of that same object” (James et al 2001 pp.118).

In the present study, we examined whether active exploration of novel objects facilitates the recognition of views of other objects in the same category (view generalization). If active exploration of object views facilitates generalization to different objects in the same category, this facilitation does not occur in the encoding process but in the matching process. Participants performed one set of generalization task of paper-clip objects (pre-observation generalization phase). Subsequently, they actively explored or passively observed views of different paper-clip objects (observation phase). After the exploration or observation, they performed two more sets of generalization tasks (1st and 2nd post-observation generalization phases). Comparison of recognition performance between the pre-observation and post-observation generalization phases permits examination of the effect of active exploration on the process of view matching.

To eliminate effects of participants' experiences with particular objects, we used novel complicated 3-D objects, termed paper-clip objects (Bülthoff and Edelman, 1992), as stimuli. Since Bülthoff and Edelman's study, paper-clip objects have been

used in a number of psychological studies to investigate object recognition. It has been shown that, even in the case of use of such complicated novel 3-D objects, human observers and primates can generalize novel views within a certain rotational range (approximately $\pm 30^\circ$) in relation to a learned view (Logothetis et al 1994). Moreover, Logothetis et al. (1995) electrophysiologically recorded responses of neurons from macaque IT and found view-specific neurons tuned to a single view of paper-clip objects. These studies provide basic behavioral and physiological data for view generalization properties of paper-clip objects. Paper-clip objects are thus appropriate stimuli for the present investigation of view generalization.

View-generalization can play an important role in achieving object constancy with a few sampled object representations. Thus, investigating the properties of change in view generalization performance after active exploration or passive observation enables examination of the effect of active exploration on the view matching process. Furthermore, we limited the range and axis of exploration in order to examine whether the effect of exploration is specific to the range and axis of exploration.

2. Experiment 1

2.1. Participants

There were fifty-four participants aged 20 to 30 years, who were unaware of the purpose of this experiment. All reported normal or corrected-to-normal vision.

2.2. Stimuli and Apparatus

Figure 1 shows an example of the stimuli used in the present experiment. All objects consisted of seven sticks. Objects were created by the following procedure. The initial point was randomly generated in a unit cube centered at the origin. In the second step, the second point was randomly generated in the unit cube at a certain distance from the first point. In the i th step, the i th point was randomly generated at a certain distance from the $(i-1)$ th point. Eight points separated by equal distances were thus randomly generated. Dodecagonal prisms centered on the line connecting two points were rendered using the Lambertian surface reflectance model. The joint parts of two adjacent dodecagonal prisms were interpolated to be smooth. We discarded objects in which any two sticks overlapped. The stimuli were presented on the center of a black background on a 21-inch CRT monitor (EIZO FlexScan T966) in a dark room. They subtended a visual angle of approximately 8° at a distance of 80 cm. We used a personal computer (DELL Dimension 8300) to present stimuli and collect participant responses.

[Figure 1]

2.3. Procedure

The experiment consisted of three generalization phases (a “pre-observation generalization phase” and two “post-observation generalization phases”) and an observation phase inserted between the first and second generalization phases, i.e. pre-

and 1st post-observation generalization phases. In the generalization phase, participants performed a sequential matching task with two serially presented objects (Figure 2). On each trial in the generalization phase, the participants first fixated a cross in the center of the display for 500 ms. Following a blank for 500 ms, a target object was presented oscillating within -2° to 2° from the training view about the x- or y-axis. The order of direction of oscillation was randomized for each trial. After 1500 ms, the training view was extinguished and the display was blanked for 1000 ms. Following a fixation cross and a blank for 500 ms each, a single static view (test view) was displayed. The test view was either a view of the target or one of non-target objects (distractors). The test view for the target object was selected among 21 views of the target, which were generated by x- and y-axis rotation of the training view within the range of -75° to 75° sampled at 15° intervals (Figure 3)¹. Participants were asked to memorize the shape of the target object during the presentation of its training view and indicate whether or not the test view was a view of the target by pressing the corresponding key on a keyboard. Although test views were presented until the participants responded, they were asked to respond as quickly and correctly as possible. After a yes/no response, the participants were asked to rate their confidence in their answer on a two-point scale (confident or not so confident) by pressing the corresponding key. After confidence rating, the next trial started. One

¹ We did not use z-axis rotation views for the following reasons: plane rotation is thought to be a special case of image transformation since no visual features are occluded by this rotation and the rotational axis must be along the line of sight of the observer. Moreover, plane rotation is rarely observed in everyday situations (Lawson, 1999).

generalization phase consisted of 126 trials. We used 21 target objects and 63 distractor objects. The number of trials in which the test view was a target was equal to the number of trials in which the test view was a distractor. In each generalization phase, participants underwent three target trials for each target object in which test views were from a different viewpoint. No feedback was given regarding whether or not responses were correct.

In the observation phase, participants observed five different paper-clip objects, each of which was presented in the center of the display for 20 sec. These objects were never used as stimuli in the generalization phases. Participants were divided into three groups, ‘active’, ‘passive’, and ‘still’, of eighteen people each. By manipulating a trackball (Microsoft Trackball Explorer) using the right hand, participants in the active group were allowed to rotate the object within the range of -45° to 45° about the x-axis from its 0° view, which was arbitrarily chosen for each object. Each participant in the passive group observed an image sequence of the object actively explored by each corresponding participant in the active group. Participants in the still group observed a static view of the object. The participants in the active and passive groups were instructed to observe the object while paying attention to variation in the object's appearance due to vertical rotation. All participants were told that they did not need to memorize the shape of the object. Note again that none of the objects in the observation phase were used in the generalization phases. Therefore, even if particular views of the objects happened to be memorized during the observation phase,

they would not be used to perform subsequent generalization phases. The order of five objects was randomized for each participant.

Participants first completed the pre-observation generalization task and observation task. After 5-min rest, they performed the first and second post-observation generalization tasks. This allowed us to determine how the effects of active exploration or passive observation change over time. At the beginning of the experiment, the participants performed 20 practice trials. The stimuli were paper-clip objects that never appeared across all experimental phases. During the practice trials, a beep sound was presented as feedback when participants made an incorrect response. The entire experiment took approximately 90 minutes to complete (see figure 3).

[Figure 2]

[Figure 3]

2.4. Data analysis

We conducted ROC analysis of pooled confidence rating data from each group of participants. The software RscorePlus (Harvey, 2007) was used to calculate a maximum likelihood estimate of the area under the ROC and its standard error. In this paper, we will use the term “true hit rate”² (Logothetis et al., 1994) to denote the area measure. The true hit rate is an index of criterion-independent accuracy of participants’ performance. In the statistical test for the effect of experimental phases,

² We were motivated to use true hit rates as an index of accuracy in order to compare our results with those of Logothetis et al (1994). Note that our calculation method of true hit rates was strictly different from theirs because they did not use rating data.

we performed a one-way ANOVA on the true hit rates for each view of each group. The within-condition variances were derived from the squares of the standard error of the estimated true hit rate.

In addition, in order to quantitatively evaluate the range of view generalization, we fitted a modified Gaussian function to the data of true hit rates of each group. The modified Gaussian function follows the form of a psychometric function with upper and lower bounds:

$$G(\theta; \sigma, b, c) = b + (1 - b - c) \exp\left(\frac{-\theta^2}{2\sigma^2}\right),$$

where c is the upper bound and b is the lower bound. The range of view generalization is defined by the value of σ (sigma). This function assumes that the data are symmetric with respect to the direction of rotation, but preliminary inspection and fitting of the data suggested that in some cases, the range of view generalization should differ between plus and minus directions of rotation. Therefore, we additionally fitted another modified Gaussian function that takes into account such asymmetry:

$$G_{asymmetric}(\theta; \sigma_1, \sigma_2, b, c) = \begin{cases} 1 - c & \theta = 0 \\ b + (1 - b - c) \exp\left(\frac{-\theta^2}{2\sigma_1^2}\right) & \theta < 0 \\ b + (1 - b - c) \exp\left(\frac{-\theta^2}{2\sigma_2^2}\right) & \theta > 0 \end{cases},$$

where σ_1 defines the range for the minus direction of rotation and σ_2 for the plus

direction. These symmetric and asymmetric models were fitted to the data of each group by least squares, under the constraint that the upper bound c is common between the axes of rotation (x and y). Since the symmetric model is a nested model of the asymmetric one, we used an F-test to determine which model better fitted the data. A statistical difference of the sigmas across the experimental phases was evaluated by a one-way ANOVA with the within-condition variances derived using the asymptotic variances of the estimated sigmas.

2.5. Results

The true hit rates for each view are shown in Figure 4. In the active group, we found significant differences among three phases in the true hit rate for $\pm 45^\circ$ x-axis rotation views (-45° : $F(2, \infty) = 4.12$, $p < 0.05$; 45° : $F(2, \infty) = 3.50$, $p < 0.05$). Post hoc comparison using the z test with Bonferroni correction revealed significant differences in the true hit rate between the pre-observation and 1st post-observation generalization phases ($p < 0.05$), and between the pre-observation and 2nd post-observation generalization phases ($p < 0.05$) for the -45° view. For the 45° view, there was a significant difference only between the pre and 2nd post-observation generalization phases. There were also significant increases for -60° and 45° y-axis rotation views in the active group (-60° : $F(2, \infty) = 6.17$, $p < 0.01$; 45° : $F(2, \infty) = 8.25$, $p < 0.01$). Post hoc comparison revealed significant differences between the pre- and 2nd post-observation generalization phases for both views ($p < 0.05$). In the passive group,

there was no significant increase in the true hit rate for x-axis rotation views. However, there was a significant increase for the -60° y-axis rotation view ($F(2,\infty) = 8.24$, $p < 0.01$). Post hoc comparison revealed significant differences between the 1st and 2nd post-observation generalization phases. In the still group, there were significant increases for the -60° and 30° x-axis rotation views (-60° : $F(2,\infty) = 3.50$, $p < 0.05$; 30° : $F(2,\infty) = 4.81$, $p < 0.01$). Post hoc comparison revealed significant differences between the pre and 2nd post-observation generalization phases for both views. There was also a significant increase for the 30° y-axis rotation view ($F(2,\infty) = 8.97$, $p < 0.01$). Post hoc comparison revealed significant differences between the pre-observation and 1st post-observation generalization phases ($p < 0.05$), and between the pre-observation and 2nd post-observation generalization phases ($p < 0.05$).

The results of Gaussian fitting to the true hit rates are also plotted in Figure 4. The symmetric model was better fitted to the data for the active group. For the other groups, the asymmetric model was better fitted. The estimated values of sigma are shown in Table 1. The ANOVA on the sigmas revealed a significant effect of generalization phases only for the x-axis rotation in the active group ($F(2,\infty) = 4.14$, $p < 0.05$).

A surface map illustrating the relative time that the participants in the active group spent at each view is shown in Figure 5, which indicates that all of the participants spent more time exploring the $\pm 45^\circ$ views than the intermediate views.

[Figure 4]

[Figure 5]

[Table 1]

2.6. Discussion

Significant effects on the true hit rate for $\pm 45^\circ$ x-axis rotation views were observed in the active group, and not in the passive group, even though participants in both groups saw the same visual sequence during the observation task. This result strongly suggests that active exploration expanded the view generalization range about the x-axis. There were also two x-axis rotation views in the still group for which there was significant improvement in the post-observation generalization phases. However, comparison of the profiles of the fitted curves shows a difference in the improvement of performances between the active and still groups. In the active group, the width of the curve becomes broader in the later generalization phases, whereas in the still group no such tendency is found. This difference is evident in the sigmas of those curves. Figure 6 shows the estimated values of sigma for all groups as a function of the order of the generalization phases. It is clear that in the case of x-axis rotation, an increase in the size of sigma was most marked and statistically significant in the active group. This provides further support to the view that active exploration can expand the range of view generalization.

In y-axis rotation views, there was a significant improvement of generalization for two views in the active group. This might indicate transfer of the effect of active

exploration in x-axis rotation to generalization of y-axis rotation views. As shown in Figure 4, however, expansion of generalization was smaller than that for x-axis rotation views in the active group. Therefore, although active exploration has some effects on the generalization of orthogonal rotation views, these effects are limited. Furthermore, some significant increases in generalization performance were observed in the passive and still groups as well. It is possible that the improvement of generalization observed in the y-axis rotation views was an effect of practice. This possibility is further suggested by the finding that the estimated values of sigma for all conditions for y-axis rotation views exhibited similar growth profiles (Figure 6). This suggests that the view transformation processes for x-axis and y-axis rotations are independent of each other. In daily life, we experience horizontal or vertical changes in viewpoint with respect to objects according to our movements. These horizontal and vertical changes are caused by different types of movements, for instance, walking around the object, and looking up or down at the object, respectively. Thus, it is plausible that these two processes work independently.

[Figure 6]

3. Experiment 2

In the active group, active exploration in the range of -45° to 45° about the x-axis resulted in significant effects on the true hit rate for $\pm 45^{\circ}$ x-axis rotation views, but not for views more than 60° apart from the training view. This finding can be interpreted

in two ways: 1) the improvement of view generalization is dependent on the range of active exploration or 2) regardless of the range of active exploration, it is simply difficult to generalize views beyond 60° . To examine these possibilities, we set a new condition for the active group in which the possible range of rotation was -30° to 30° . If, regardless of the range of active exploration, active exploration facilitates view generalization within the range of -45° to 45° , then active exploration of views in the range of -30° to 30° should yield the same results as obtained in experiment 1.

3.1. Procedure

The experimental procedure and apparatus were the same as for the active group in experiment 1, except for the range of active exploration. There were 18 participants, aged 20 to 26 years, who were unaware of the purpose of this experiment.

3.2. Results

The true hit rates for each view and the results of Gaussian fitting are shown in Figure 7. There was no significant increase in the true hit rate for any rotation view. The asymmetric model was better fitted to the data for the x-axis rotation. For the y-axis rotation, the symmetric model was better fitted. The estimated values of sigma are shown in Table 2. The ANOVA on the sigmas revealed no significant effect of the generalization phases for either the x-axis rotation or the y-axis rotation.

[Table 2]

[Figure 7]

3.3. Discussion

Although the true hit rates for $\pm 30^\circ$ x-axis rotation views exhibited a slight increase in later generalization phases, there was no significant increase in those for any view. One possible interpretation of this result is that generalization of $\pm 30^\circ$ views is within the range of the default view generalization. Logothetis et al (1994) showed that human observers and monkeys could generalize novel views in the range of approximately 30° from one learned view. Moreover, they found view-tuned cells in the macaque IT cortex, and reported that the mean standard deviation of the tuning curve of these cells was approximately 30° (Logothetis et al 1995). This value of 30° is consistent with the results reported by psychophysical studies using wire-like objects (Rock and DiVita, 1987) and paper-clip objects (Bülthoff and Edelman, 1992; Edelman and Bülthoff, 1992). Also, in computer simulation using a GRBF network performed by Poggio and Edelman (1990), when the network could discriminate all views of a target from distractors, the range for one GRBF unit in the hidden layer was approximately 30° . In experiment 2, since $\pm 30^\circ$ views were within the range of the default view generalization, recognition performances for these views were already relatively high in the pre-observation phase. Therefore, the effect of active exploration might be weaker for $\pm 30^\circ$ than that for $\pm 45^\circ$ views. Nevertheless, active exploration in the range of -30° to 30° did not improve generalization performance for

$\pm 45^\circ$ views. This supports the hypothesis that improvement of view generalization depends on the range of active exploration. Moreover, there was no improvement for y-axis rotation views, indicating that active exploration in the range of -30° to 30° about the x-axis did not facilitate generalization of orthogonal axis rotation views or expand the generalization range. This supports the results of experiment 1.

4. Experiment 3

In the passive group in experiment 1, no significant improvement was observed in any view. However, the participants could have had difficulty in predicting the movement of objects since it was generated by other participants. This lack of predictability of subsequent movement may inhibit facilitation of view transformation. In order to test this possibility, we set a new condition for the passive group in which objects rotated at a constant speed in the range of -45° to 45° . If ease of prediction of the next change in views is important for improvement of view generalization, participants who observe objects rotating at a constant speed should exhibit improvement of performance of view generalization.

4.1. Procedure

The experimental procedure and apparatus were the same as in experiment 1 for the passive group, except that the rotation speed was constant. The rotation speed was adjusted so as to present the view sequence, 0° to 45° (-45°) to 0° to -45° (45°) to 0° ,

twice for 20 seconds. This speed was based on typical data for active exploration obtained for the participants in the active group in experiment 1 (Figure 8). The order of direction of rotation was randomized for each object. There were 18 participants, aged 21 to 29 years, who were unaware of the purpose of this experiment.

[Figure 8]

4.2. Results

The true hit rates for each view and the results of Gaussian fitting are shown in Figure 9. There was no significant increase in the true hit rate for any rotation view. The asymmetric model was better fitted for both the x-axis rotation and the y-axis rotation. The estimated values of sigma are shown in Table 3. The ANOVA on the sigmas revealed no significant effect of the generalization phases for either the x-axis rotation or the y-axis rotation.

[Table 3]

[Figure 9]

4.3. Discussion

The results indicate that ease of prediction of the next view does not lead to improvement of view generalization performance in passive observation. Accordingly, we argue that it is necessary to observe manually generated changes in views for improvement of view generalization performance.

5. Experiment 4

In experiment 1, we showed that active exploration of views around the x-axis facilitates generalization performance on x-axis rotation views. However, it is possible that this effect is idiosyncratic for x-axis rotation. In experiment 4, we examined whether active exploration of views around the y-axis facilitates generalization of y-axis rotation views.

5.1. Procedure

The experimental procedure was the same as in experiment 1 for the active and passive groups, except that the range of active exploration or passive observation was -45° to 45° about the y-axis. In the observation phase, the participants in the active group rotated the objects used in the observation phase in experiment 1 by rotating a turntable using their right hand. The rotation angle was detected by using a 3-D motion sensor (developed by NEC/Tokin) attached on the center of the turntable. This sensor can provide its 3-D attitude information by detecting geomagnetism. Participants in the passive group observed an image sequence of the object actively explored by each corresponding participant in the active group. There were 18 participants for each group, aged 21 to 29 years, who were unaware of the purpose of this experiment.

5.2. Results

The true hit rates for each view and the results of Gaussian fitting are shown in Figure 10. In the active group, there were significant increases in true hit rate for -45° ($F(2,\infty) = 3.66$, $p < 0.05$), and -30° ($F(2,\infty) = 3.08$, $p < 0.05$) y-axis rotation view, and 30° x-axis rotation view ($F(2,\infty) = 3.92$, $p < 0.05$). Post hoc comparison using the z test with Bonferroni correction revealed significant differences in the true hit rate between the pre-observation and 1st post-observation generalization phases ($p < 0.05$) for -45° y-axis rotation view and 30° x-axis rotation view. However, for -30° y-axis rotation view, post hoc test showed no significant differences at $p < 0.05$. In the passive group, there was no significant increase in true hit rate for any view. The estimated values of sigma are shown in Table 4. The asymmetric model was better fitted to the data for y-axis rotation in the active group and that for x-axis rotation in the passive group. The ANOVA on the sigmas revealed a significant effect of generalization phases only in the minus direction of y-axis rotation in the active group ($F(2,\infty) = 3.53$, $p < 0.05$).

[Table 4]

[Figure 10]

[Figure 11]

5.3. Discussion

The results demonstrate that active exploration of views of objects about the y-axis

facilitated view generalization. Moreover, we found that the facilitation in view generalization was specific to the minus direction of rotation about the y-axis. A surface map illustrating the relative time that the participants in the active group spent at each view is shown in Figure 11, which indicates that all of the participants spent more time exploring the $\pm 45^\circ$ views than the intermediate views. In order to examine whether there was a significant difference in times for which the participants observed views in the minus and plus direction of rotation about the y-axis, we conducted a paired t-test for relative times spent at $-45^\circ \sim -35^\circ$ and $45^\circ \sim 35^\circ$ y-axis rotation views. This analysis revealed that the participants spent significantly longer time at $-45^\circ \sim -35^\circ$ views than $45^\circ \sim 35^\circ$ views ($t(17)=5.16$, $p<0.001$, two-tailed test). On the other hand, the same analysis for the data of active exploration in experiment 1 did not show a significant difference ($t(15)=0.998$, n.s.³). Therefore, it may be particular to the y-axis rotation that the participants tended to explore views in the minus direction of rotation. There is, however, another possibility that this difference depended on the device to rotate objects in active exploration. These possibilities should be investigated in the future study.

6. General Discussion

The present findings suggest that active exploration can facilitate view-generalization, supporting the results of Harman et al (1999) and James et al

³ In experiment 1, we could not collect the data of active exploration from 2 participants because of problems of the program.

(2001). James et al (2001) has already reported that active exploration facilitates matching between two different views. As noted above, however, since they used the same set of stimuli in both the exploration phase and matching experiment, the possibility remained that particular views were explicitly memorized and used for matching. In the present experiments, we found that active exploration of views of novel objects facilitated view generalization of other objects of the same category. Accordingly, we argue that the improvement observed in the active group in experiment 1 and 4 was not due to facilitation of the encoding process, but to the matching process.

Which object recognition process was facilitated by active exploration? One possibility is mental rotation (Shepard and Metzler 1971). The present findings can be explained by the multiple views plus transformation theory proposed by Tarr et al (e.g., Tarr 1995; Tarr and Pinker 1989). Since the degree of change in visual features in the rotational range of $\pm 30^\circ$ from one view is relatively small, views in the range of -30° to 30° about the training view can be directly recognized by a default view generalization process such as template matching based on a view-tuned cell. However, rotation beyond $\pm 30^\circ$ greatly distorts visual features and results in qualitative changes in views because of self-occlusion. Template matching can be no longer used to generalize such views. Thus, it is necessary to transform visual features three-dimensionally. As the angular difference between the training view and the test view grows larger, such transformation is more time-consuming and results in a higher

error rate. Accordingly, the improvement of generalization performance observed in experiment 1 and 4 can be interpreted as reflecting facilitation of a transformation process that is similar to mental rotation. However, it has been pointed out that view-dependent object recognition and mental rotation are based on different neural substrates (Gauthier et al., 2002). Gauthier et al. (2002) found that activity in the superior parietal lobe increased depending upon the angular disparity during mental rotation, but not during object recognition. This dissociation implies that transformations such as mental rotation are not recruited in object recognition. However, they also found regions (SPL/IPL, precentral gyrus) where activity increased view-dependently both in mental rotation and in object recognition tasks. It thus remains unclear whether mental rotation is actually involved in object recognition.

It is also possible that improvement of view generalization is due to an effect of practice. In the present experiments, however, any such effect would have been equal for all groups. The improvement observed in the active group is thus unlikely to be the result of practice alone. Nevertheless, it is likely that there was an interaction between the factors of practice and active exploration. In experiment 1, the true hit rate for 45° x-axis rotation view did not increase immediately after active exploration, but increased gradually as the number of generalization phase increased. On the other hand, the true hit rate for -45° x-axis rotation view reliably increased even in the first post-observation generalization phase. These results indicate an asymmetric facilitation of view generalization with respect to the direction of rotation, suggesting

the existence of interaction between the effects of practice and active exploration. In experiment 4, such asymmetric facilitation was found in the y-axis rotation views. One possible interpretation of these findings is that the asymmetry was caused by an anatomical constraint of the hand. In the active group of experiment 1, more facilitation was observed for the minus direction of x-axis rotation. This direction corresponds to clockwise rotation against the right hand manipulating an object from the right side. Similarly, the facilitative direction observed in experiment 4 corresponds to clockwise rotation against the right hand manipulating an object from above. We conjecture that an anatomical constraint of the right hand results in a preference in the direction of the hand rotation, which gives rise to asymmetric facilitation of view generalization in the corresponding direction of object rotation.

What did participants learn through active exploration? One possibility is that participants acquired skill of mental rotation through interactions between the motor and visual systems. Recent behavioral and brain studies suggest that the motor system is involved in mental rotation, reporting activation in motor areas, such as premotor or motor cortex, during mental rotation tasks (e.g., Cohen et al 1996; Richter et al 1997, 2000). This suggests that motor imagery is used in the mental rotation task. Thus, active exploration may provide an efference copy and/or proprioceptive information that facilitates the process of mental rotation. Such a relationship between motor and mental rotation has also been suggested by the results of behavioral studies (Wohlschläger and Wohlschläger 1998; Wexler et al 1998). For instance,

Wexler et al (1998) reported that manual physical rotation could facilitate mental rotation. They showed that when participants rotated a joystick manually, the reaction time for mental rotation to the direction compatible with rotation of the joystick was faster than that in the incompatible direction. This suggests that some processes are common to motor and mental rotation. In the present study, the results of experiment 2 showed that the expansion of view generalization range was limited to the range of exploration. This suggests that it is necessary not only to rotate objects manually but also to observe how views of objects vary as the result of rotation for improvement of view generalization. The results of experiment 3 also suggest that interaction between motor and visual information is necessary for improvement of view generalization. Taken together, the present findings suggest the contribution of motor information to achievement of object constancy.

Another possibility is that participants learned a general view-transformation rule regarding paper-clip objects through active exploration. In daily life, a novel view of a car can be recognized even when the car is seen for the first time. This suggests that a general view-transformation rule is used, permitting generalization to a novel view of a novel car. Similarly, in the present experiment, participants could generalize a novel view of a novel paper-clip object using a general view-transformation rule regarding paper-clip objects. This appears to be similar to the “expertise effect” found by Gauthier et al (1999), who showed that after intensive training in discriminating novel objects in a category (‘greebles’), activation in the middle fusiform gyrus

corresponding to the ‘face-specific’ area increased more than in novices during a sequential matching task of greebles, suggesting that the “face-area” is recruited for general subordinate-level recognition. In their experiment, different exemplars were used in training and fMRI sessions, indicating that the expertise effect occurred not at the exemplar-level but at the within-category level. Our findings might reflect a type of expertise effect in the ventral region of the brain. However, our participants did not need to perform a cognitive task in the observation phase, and only observed how views were changing with their hand movement. Moreover, our observation phase took much less time than their training phase (“approximately 7 hours over at least 4 days (Gauthier et al, 1999, p.572)”). Thus, our expertise effect may reflect a process of facilitation different from those found by Gauthier et al. One possible interpretation of our findings is that the participants learned view-transformation rules for x-axis rotations of 45° and -45° specifically. The surface map obtained might support this speculation. Figure 5 shows that all participants tended to spend more time at the $\pm 45^\circ$ x-axis rotation views than the intermediate views. Actually, there were significant improvements only for $\pm 45^\circ$ x-axis rotation views, and not for views from -30° to 30° . Although these two possibilities are not mutually exclusive, it will be important to determine whether the improvement was due simply to facilitation of mental rotation or to acquisition of a category-specific view-transformation rule.

In the active group of experiment 1, there were two y-axis rotation views for which generalization performance was significantly improved in the post-observation

generalization phases. This improvement indicates that active exploration of x-axis rotation views affected generalization performance for the orthogonal rotation views. However, this effect was not as clear as for the x-axis rotation views, since expansion of the generalization range was not clear in the y-axis rotation views. The effect of active exploration of x-axis rotation views thus did not transfer perfectly to view-generalization of the y-axis rotation views. Figure 12 illustrates the estimated values of sigma for all groups in experiments 1-4 as a function of the number of generalization phases which participants completed. It indicates again that the active group in experiment 1 exhibits a clear increase in sigma only for the generalization of x-axis rotation views. This suggests that the view-transformation process is basically independent for each rotational axis. Furthermore, because there were also significant increases in y-axis rotation views for the passive and still groups in experiment 1, this facilitation might be due to practice, as noted above. However, Peissig et al (2002) showed that pigeons could generalize novel views on a rotational axis that was orthogonal to the axis in training with multiple views. This finding suggests the possibility that there is interaction between the view-transformation processes about the x-axis and y-axis. However, their stimuli were single 3-D volumetric parts (geons; Biederman 1987). It is thus possible that the pigeons simply responded to the novel views based on similarity to the training views rather than rotating the objects systematically.

[Figure 12]

In experiment 1, there were also significant improvements of generalization in the still group. As seen in Figure 4 and Figure 6, however, there was no statistically significant expansion of the generalization range as observed in the active group. This suggests that the improvements in the active and still groups were essentially different. The participants in the still group had the opportunity to transform objects mentally in this condition during the observation task. Such “mental” exploration might affect view-generalization performance. Nevertheless, the effect of active exploration is clearer than the effect of observation of a still view.

Previous object recognition models were based solely on visual processing (e.g., Poggio and Edelman 1990; Riesenhuber and Poggio 1999). However, the present findings suggest that visuo-motor interaction can affect object recognition and contribute to object constancy. Wexler et al (1998) pointed out that visuo-motor interaction would take place in the premotor and posterior parietal cortex. These brain areas have been shown to be important for recognizing objects from non-canonical viewpoints. Warrington and Taylor (1973) showed that lesions in right parietal cortex lead to deficits in recognizing common objects from unusual viewpoints (non-canonical views). Sugio et al (1999) showed that the right parietal lobule and ventral premotor cortex were activated when participants observed non-canonical views of common objects. These lines of evidence suggest that recognizing non-canonical views requires information from dorsal stream areas. Milner and Goodale (1995) suggested that the dorsal stream in the brain processes visual

information regarding actions such as grasping and manual control of objects. Recent brain imaging studies have suggested that the posterior parietal cortex and other dorsal visual areas are involved in performing visual tasks that require a precise analysis of metric visual information (James et al 2002, 2003). Precise and metric visual information is important for visually guided action, since metric information such as the direction of the main axis of objects, the width and length of objects, and the spatial relationships among visual features is important for grasping. Furthermore, these pieces of information can comprise a stable representation of objects, in that they are maintained with respect to large changes of viewpoints. We speculate that active movement facilitates the retrieval of such stable visual information of objects through a premotor-parietal network, and that this facilitation enables improvement of mental rotation ability or the acquisition of a view-transformation rule for object recognition from various viewpoints. In the literature on object recognition research, there has been debate over whether the essential property of object representations is qualitative or metric (e.g., Hayward and Tarr 1997, Biederman and Bar 1999). Presumably, metric and qualitative properties of 3-D objects are processed separately in the brain depending on task demand. Each theorist may emphasize only one aspect of object recognition. The need for integration of such theories has been stressed in a number of psychological studies (e.g., Foster and Gilson 2002; Hayward 2003). Brain imaging studies will further help to construct such a hybrid object recognition model.

In summary, the findings of the present study revealed that active exploration of

views of 3-D objects facilitates view generalization. This suggests that the motor system is involved in the process of recognition of 3-D objects, leading to the need for a new framework for object recognition that takes advantage of the interaction between motor and visual information.

Acknowledgments

We thank the two reviewers for their helpful comments. This research was supported by the Academic Frontier Project from the Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT). This research was also supported in part by Grants-in-Aid for Scientific Research (S) (No. 20220003) from the Japan Society for the Promotion of Science and for Young Scientists (B) (Nos 16700251 and 21700287) from MEXT.

References

- Biederman I, 1987 “Recognition-by-components: a theory of human image understanding” *Psychological Review* **94** 115-147
- Biederman I, Bar M, 1999 “One-shot viewpoint invariance in matching novel objects” *Vision Research* **39** 2885-2899
- Bülthoff H H, Edelman S, 1992 “Psychological support for a 2D interpolation theory of object recognition” *Proceedings of the National Academy of Science* **89** 60-64
- Christou C G, Bülthoff H H, 1999 “View dependence in scene recognition after active

- learning” *Memory & Cognition* **27** 996-1007
- Cohen M S, Kosslyn S M, Breiter H C, DiGirolamo W L, Thompson W L, Anderson A K, Bookheimer S Y, Rosen B R, Belliveau J W, 1996 “Changes in cortical activity during mental rotation. A mapping study using functional MRI” *Brain* **119** 89-100
- Edelman S, Bühlhoff H H, 1992 “Orientation dependence in the recognition of familiar and novel views of 3D objects” *Vision Research* **32** 2385-2400
- Foster D H, Gilson S J, 2002 “Recognizing novel three-dimensional objects by summing signals from parts and views” *Proceedings of the Royal Society B: Biological Sciences* **269** 1939-1947
- Gauthier I, Hayward W G, Tarr M J, Anderson A W, Skudlarski P, Gore J C, 2002 “BOLD activity during mental rotation and viewpoint-dependent object recognition” *Neuron* **34** 161-171
- Gauthier I, Tarr M J, Anderson A W, Skudlarski, P, Gore J C, 1999 “Activation of the middle fusiform ‘face area’ increases with expertise in recognizing novel objects” *Nature Neuroscience* **2** 568-573
- Gibson J J, 1979 *The Ecological Approach to Perception* (Boston: Houghton-Mifflin)
- Harman K L, Humphrey G K, Goodale M A, 1999 “Active manual control of object views facilitates visual recognition” *Current Biology* **9** 1315-1318
- Harvey L O, 2007 “Parameter estimation of signal detection models: RscorePlus user’s manual (Version 5.5.8)” [Computer software and manual]. Retrieved September 25, 2007, from <http://psych.colorado.edu/~lharvey/>

- Hayward W G, Tarr M J, 1997 “Testing conditions for viewpoint invariance in object recognition” *Journal of Experimental Psychology: Human Perception and Performance* **23** 1511-1521
- Hayward W G, 2003 “After the viewpoint debate: where next in object recognition?” *Trends in Cognitive Sciences* **7** 425-427
- Held R., Hein A, 1963 “Movement-produced stimulation in the development of visually guided behaviour” *Journal of Comparative & Physiological Psychology* **56** 872-876
- James K H, Humphrey G K, Goodale M A, 2001 “Manipulating and recognizing virtual objects: where the action is” *Canadian Journal of Experimental Psychology* **55** 113-122
- James T W, Humphrey G K, Gati J S, Menon R S, Goodale M A, 2002 “Differential effects of viewpoint on object-driven activation in dorsal and ventral streams” *Neuron* **35** 793-801
- James T W, Culham J, Humphrey G K, Milner A D, Goodale M A, 2003 “Ventral occipital lesions impair object recognition but not object-directed grasping: an fMRI study” *Brain* **126** 2463-2475
- Lawson R, 1999 “Achieving visual object constancy across plane rotation and depth rotation” *Acta Psychologica* **102** 221-245
- Logothetis N K, Pauls J, Bülthoff H H, Poggio T, 1994 “View-dependent object recognition by monkeys” *Current Biology* **4** 401-414

- Logothetis N K, Pauls J, Poggio T, 1995 “Shape representation in the inferior temporal cortex of monkeys” *Current Biology* **5** 552-563
- Milner A D, Goodale M A, 1995 *The visual brain in action* (Oxford: Oxford University Press)
- Peissig J, Wasserman E A, Young M E, Biederman I, 2002 “Learning an object from multiple views enhances its recognition in an orthogonal rotational axis in pigeons” *Vision Research* **42** 2051-2062
- Piaget J, Inhelder B, 1948 *The Child’s Conception of Space* (London: Routledge)
- Poggio T, Edelman S, 1990 “A network that learns to recognize three-dimensional objects” *Nature* **343** 263-266
- Richter W, Ugurbil K, Georgopoulos A, Kim S G, 1997 “Time-resolved fMRI of mental rotation” *NeuroReport* **8** 3697-3702
- Richter W, Somorjai R, Summers R, Jarmasz M, Menon R S, Gati J S, Georgopoulos A P, Tegeler C, Ugurbil K, Kim S G, 2000 “Motor area activity during mental rotation studied by time-resolved single-trial fMRI” *Journal of Cognitive Neuroscience* **12** 310-320
- Riesenhuber M, Poggio T, 1999 “Hierarchical models of object recognition in cortex” *Nature Neuroscience* **2** 1019-1025
- Rock I, DiVita J, 1987 “A case of viewer-centered object perception” *Cognitive Psychology* **19** 280-293
- Shepard R N, Metzler J, 1971 “Mental rotation of three-dimensional objects” *Science*

171 701-703

Sugio T, Inui T, Matsuo K, Matsuzawa M, Glover G H, Nakai T, 1999 “The role of the posterior parietal cortex in human object recognition: a functional magnetic resonance imaging study” *Neuroscience Letters* **276** 45-48

Tarr M J, Pinker S, 1989 “Mental rotation and orientation dependence in shape recognition” *Cognitive Psychology* **21** 233-282

Tarr M J, 1995 “Rotating objects to recognize them: a case study of the role of viewpoint dependency in the recognition of three-dimensional objects” *Psychonomic Bulletin and Review* **2** 55-82

Tong F H, Marlin S G, Frost B J, 1995 ”Cognition map formation in a three-dimensional visual virtual world” Poster presented at the IRIS/PRECARN Workshop, Vancouver, BC.

Warrington E K, Taylor A M, 1973 “The contribution of the right parietal lobe to object recognition” *Cortex* **9** 152-164

Wexler M, Kosslyn S M, Berthoz A, 1998 “Motor processes in mental rotation” *Cognition* **68** 77-94

Wohlschläger A, Wohlschläger A, 1998 “Mental and manual rotation” *Journal of Experimental Psychology: Human Perception and Performance* **24** 397-412

Table 1. Estimated values of sigma for each group in experiment 1

x-axis rotation

	Active	Passive		Still	
		$\theta < 0$	$\theta > 0$	$\theta < 0$	$\theta > 0$
Pre	21.31	22.80	29.50	25.16	26.81
Post1	27.43	29.05	37.19	19.19	27.54
Post2	41.95	29.00	32.50	24.98	28.10

y-axis rotation

	Active	Passive		Still	
		$\theta < 0$	$\theta > 0$	$\theta < 0$	$\theta > 0$
Pre	25.53	27.37	33.69	24.76	25.72
Post1	31.61	31.74	39.21	32.66	33.76
Post2	36.13	37.23	39.48	32.43	38.35

Table 2. Estimated values of sigma in experiment 2

	x-axis rotation		y-axis rotation
	$\theta < 0$	$\theta > 0$	
Pre	20.27	22.63	26.11
Post1	22.01	30.81	25.35
Post2	20.27	33.60	37.50

Table 3. Estimated values of sigma in experiment 3

	x-axis rotation		y-axis rotation	
	$\theta < 0$	$\theta > 0$	$\theta < 0$	$\theta > 0$
Pre	28.25	35.84	26.30	23.09
Post1	23.47	28.93	23.59	30.84
Post2	23.91	31.31	28.23	32.62

Table 4. Estimated values of sigma in experiment 4

Active group	x-axis rotation	y-axis rotation	
		$\theta < 0$	$\theta > 0$
Pre	20.97	18.45	27.80
Post1	22.71	28.35	28.84
Post2	25.08	30.95	32.76

Passive group	x-axis rotation		y-axis rotation
	$\theta < 0$	$\theta > 0$	
Pre	21.89	25.95	25.95
Post1	19.46	24.86	24.86
Post2	20.90	30.20	30.20

(Figure captions)

Fig. 1. Examples of stimuli. The center view is the arbitrarily defined 0° view of this paper-clip object. The surrounding views correspond to $\pm 45^\circ$ rotation views with respect to the x- (top and bottom) and y-axes (left and right) from the 0° view.

Fig. 2. Procedure for one trial of the generalization task. The horizontal axis represents time. Participants were serially presented with a sequence of stimuli on a black background, as shown in the top row. A test view was presented until participants responded. After confidence rating by the participant, the next trial was started, followed by a blank for 1,000 ms.

Fig. 3. Viewing sphere illustrating the relationship between test views and the range of active exploration. When a test view is a view of a target, the test view was one of 21 views sampled from the range of -75° to 75° from the 0° view (training view) with respect to the x- and y-axes. In the observation phase, participants in the active group explored views in the range of -45° to 45° , as shown in the figure. Note that none of the five objects used in the observation phase was used in the generalization phase. Therefore, even if the participants memorized particular views of the five objects, the memorized views would not be useful in subsequent generalization tasks.

Fig. 4. True hit rates and the results of fitting of Gaussian to the true hit rates for the active group (top), the passive group (middle), and the still group (bottom). True hit rates for x-axis rotation views (left) and y-axis rotation views (right). Here, 0° corresponds to the training view. Each plot represents the average of true hit rates of 12 participants for all targets. Asterisks represent views for which there was a significant difference in true hit rates ($p < 0.05$).

Fig. 5. Surface map illustrating the relative time that the participants in the active group spent at each view. Averaged for five objects. This indicates that participants spent more time at both ends of the range of active exploration than at intermediate views. The data of 16 participants is shown, because we could not collect the data of active exploration from 2 participants because of problems of the program.

Fig. 6. Estimated values of sigma for all groups as a function of generalization phase number for x-axis rotation views (left) and y-axis rotation views (right).

Fig. 7. True hit rates and the results of fitting of Gaussian to the true hit rates for x-axis rotation views (left) and y-axis rotation views (right) in experiment 2.

Fig. 8. Typical profile of active exploration by a participant. The horizontal axis represents time, and the vertical axis the angle from the 0° view about the x-axis.

These plots indicate that this participant tended to explore by reciprocating between -45° and 45° views two or three times. This tendency was observed for all participants.

Fig. 9. True hit rates and the results of fitting of Gaussian to the true hit rates for x-axis rotation views (left) and y-axis rotation views (right) in experiment 3.

Fig. 10. True hit rates and the results of fitting of Gaussian to the true hit rates for the active group (top), the passive group (bottom) in experiment 4. True hit rates for x-axis rotation views (left) and y-axis rotation views (right).

Fig. 11. Surface map illustrating the relative time that the participants in the active group spent at each view in experiment 4. Averaged for five objects. This indicates that participants spent more time at both ends of the range of active exploration than at intermediate views.

Fig. 12. Estimated values of sigma for all groups in experiments 1-4 as a function of generalization phase number for x-axis rotation views (left) and y-axis rotation views (right).

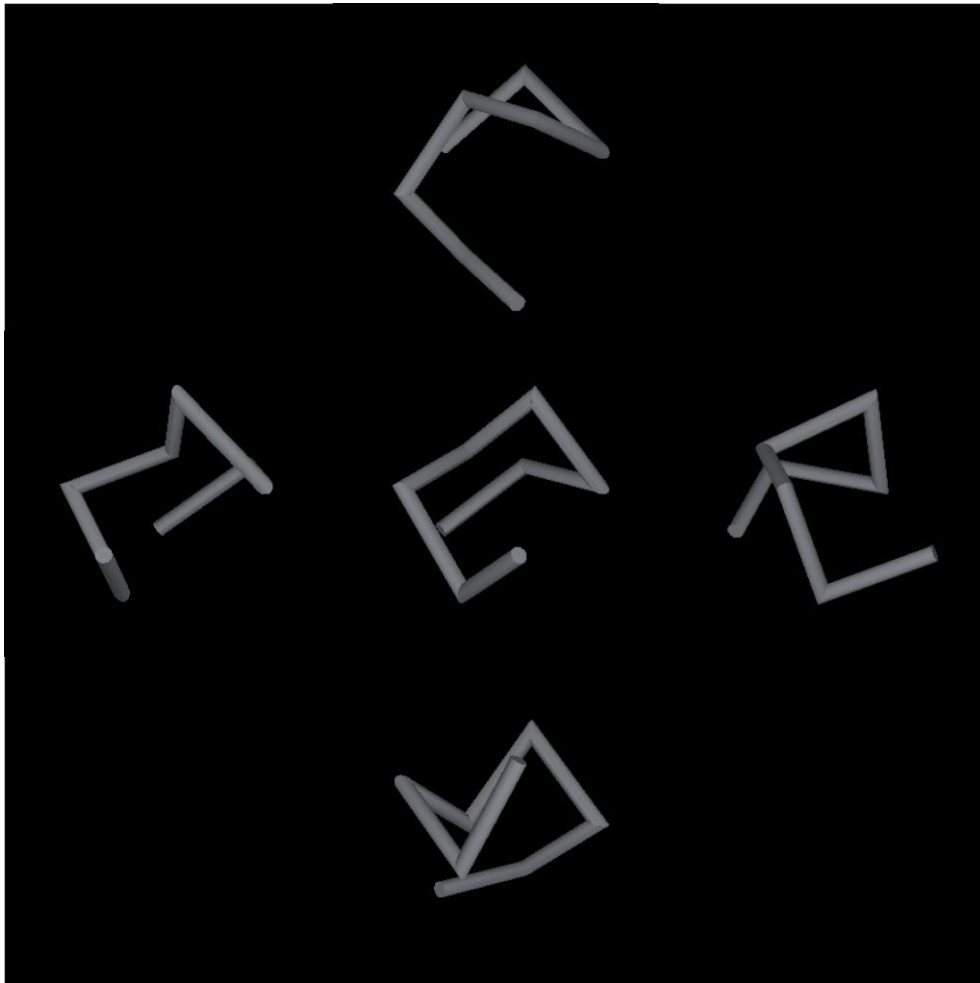


Fig. 1.

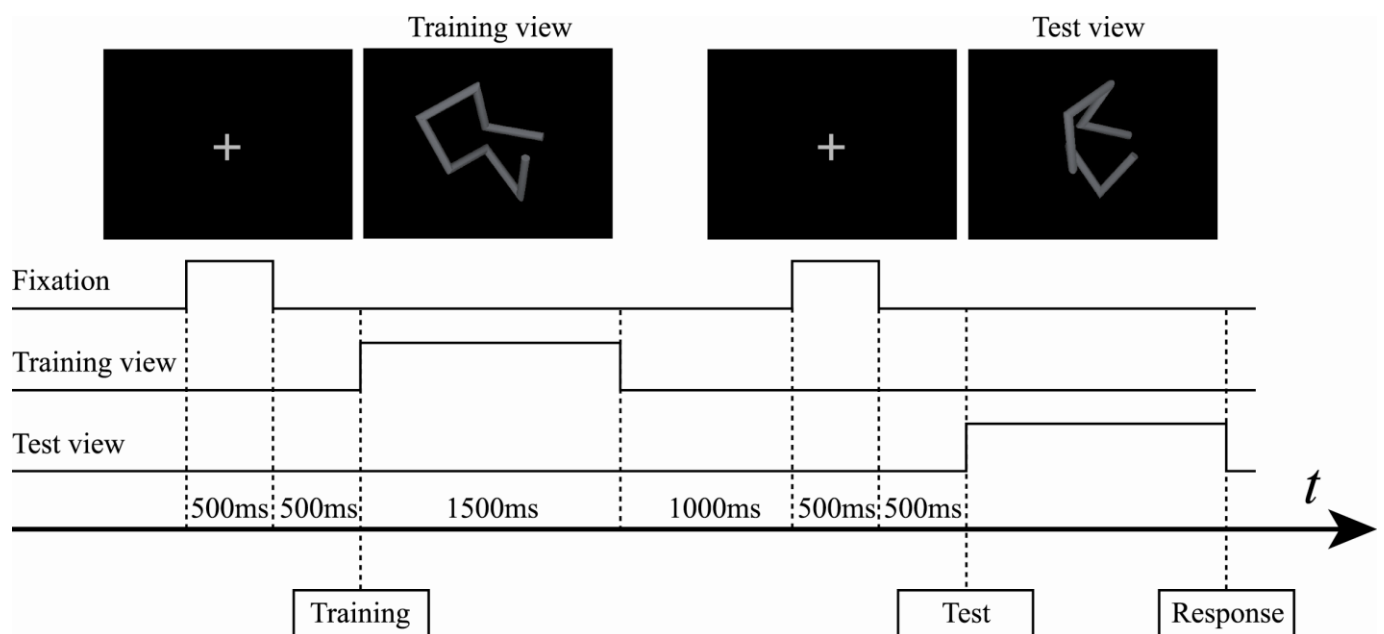


Fig. 2.

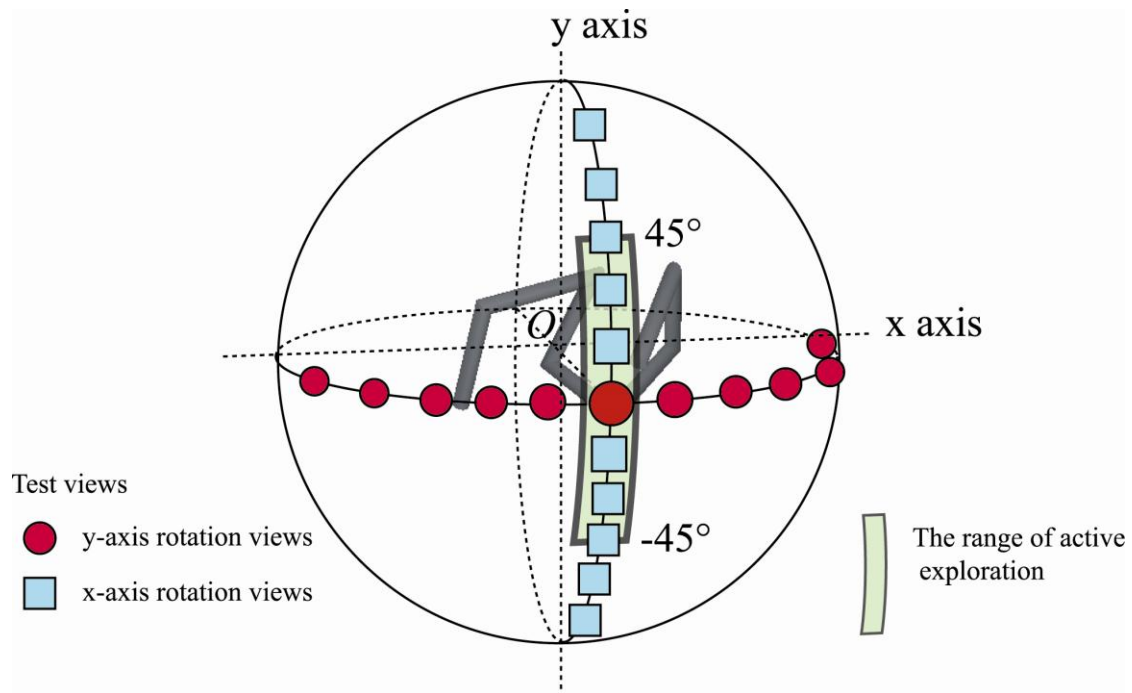


Fig. 3.

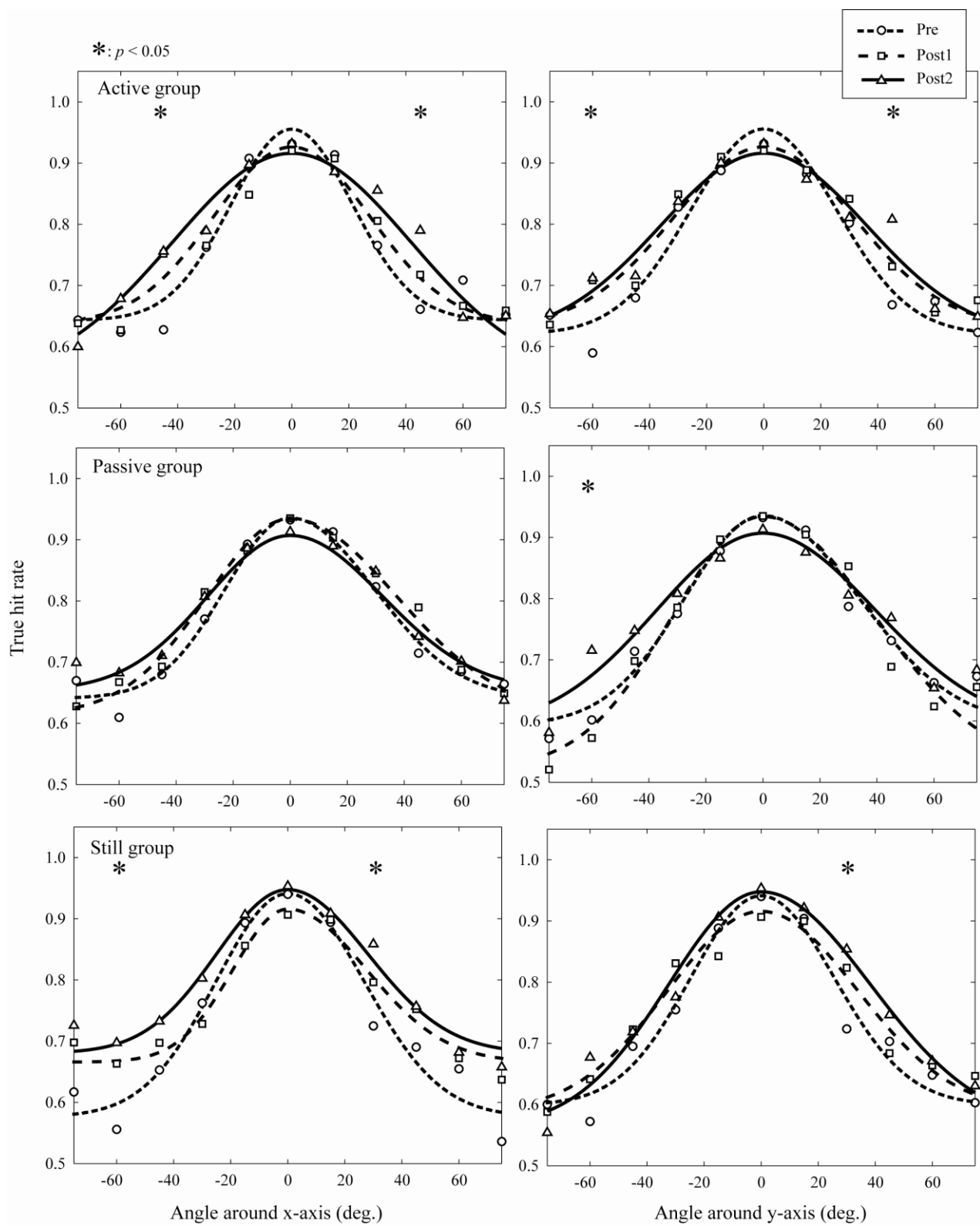


Fig. 4.

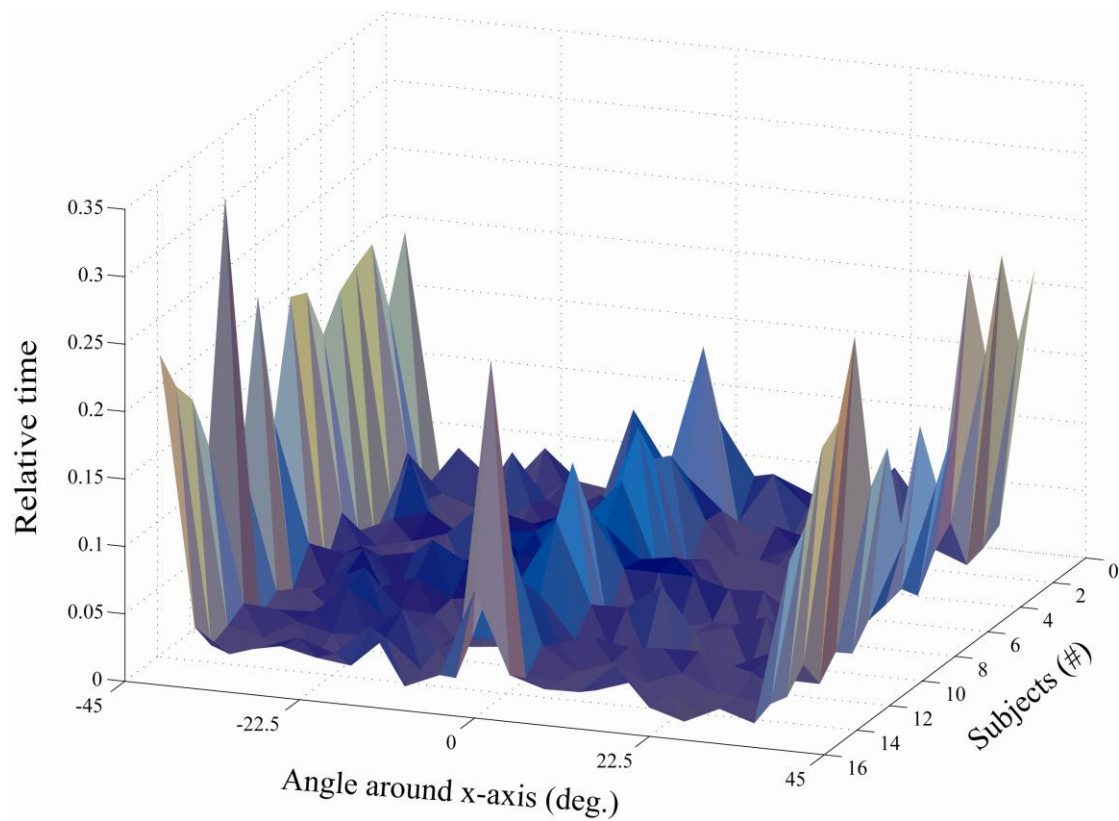


Fig. 5.

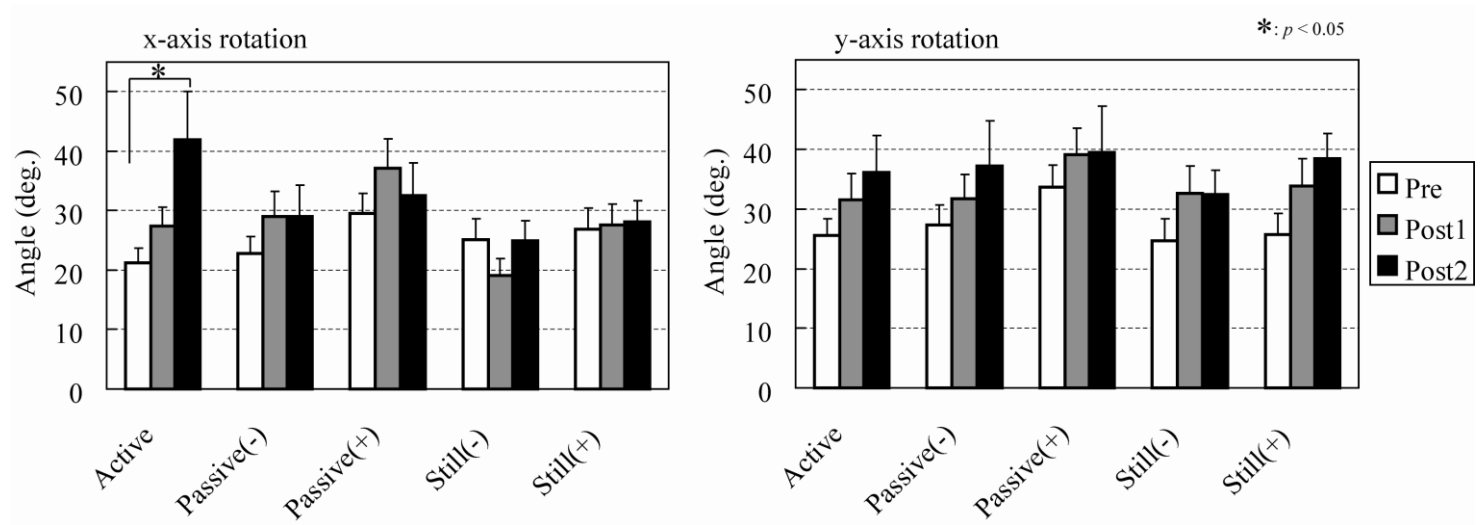


Fig. 6.

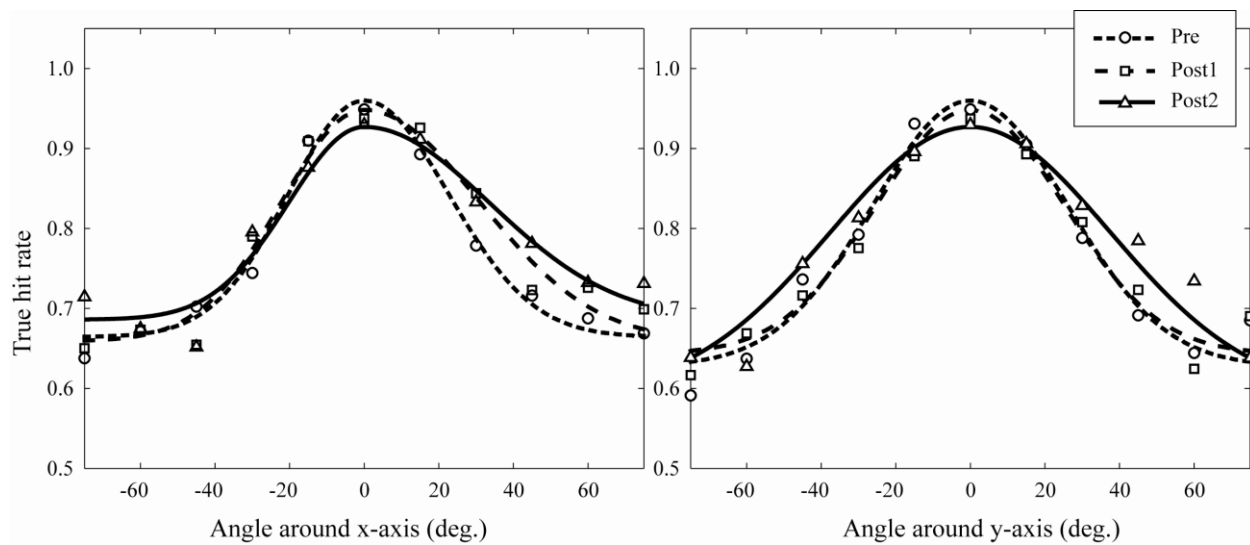


Fig. 7.

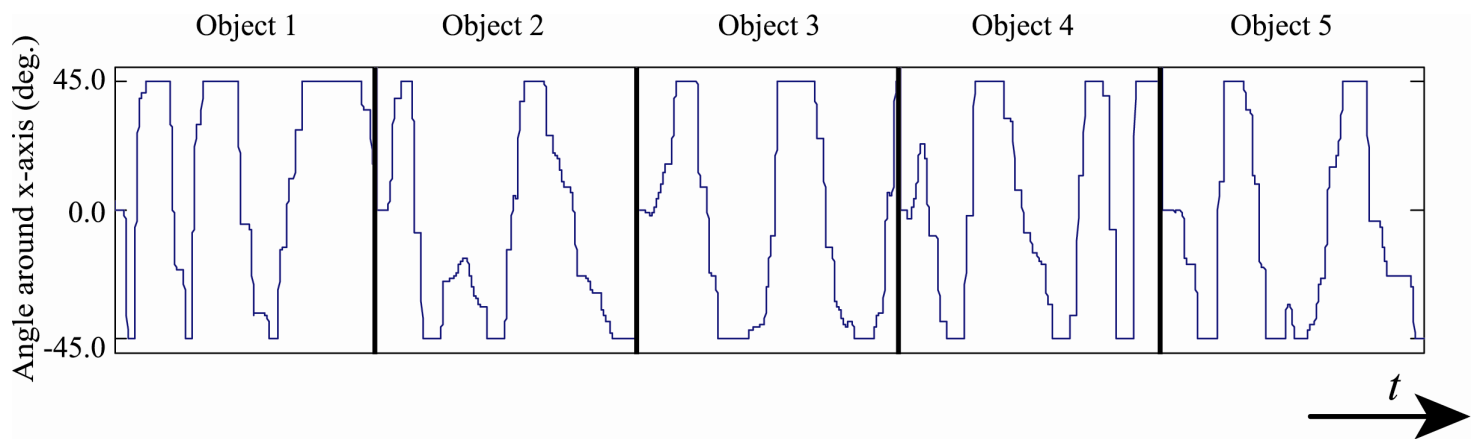


Fig. 8.

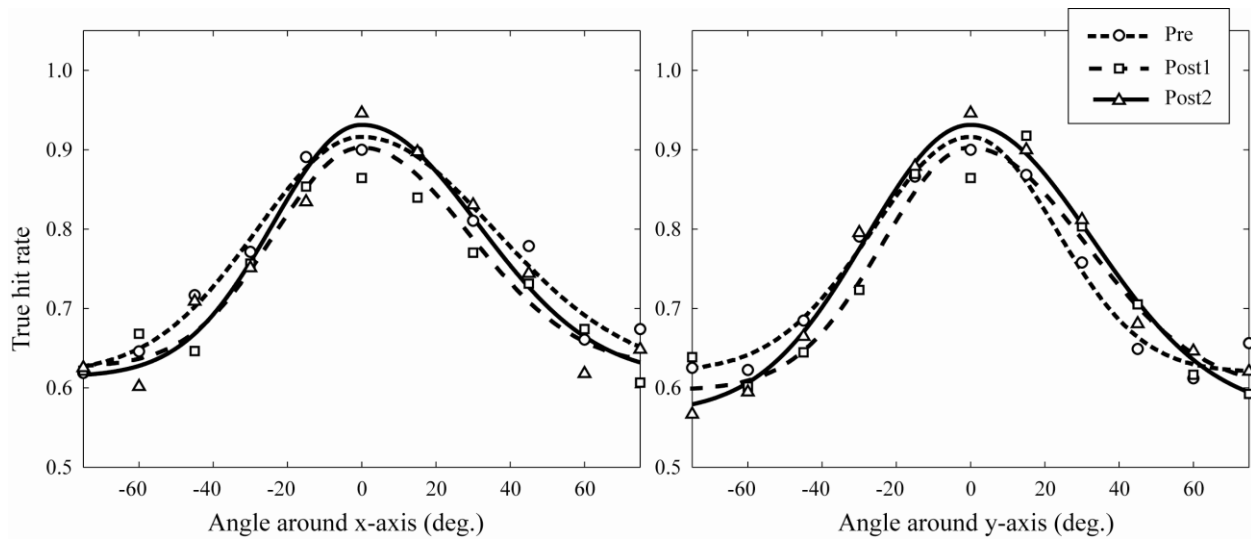


Fig. 9.

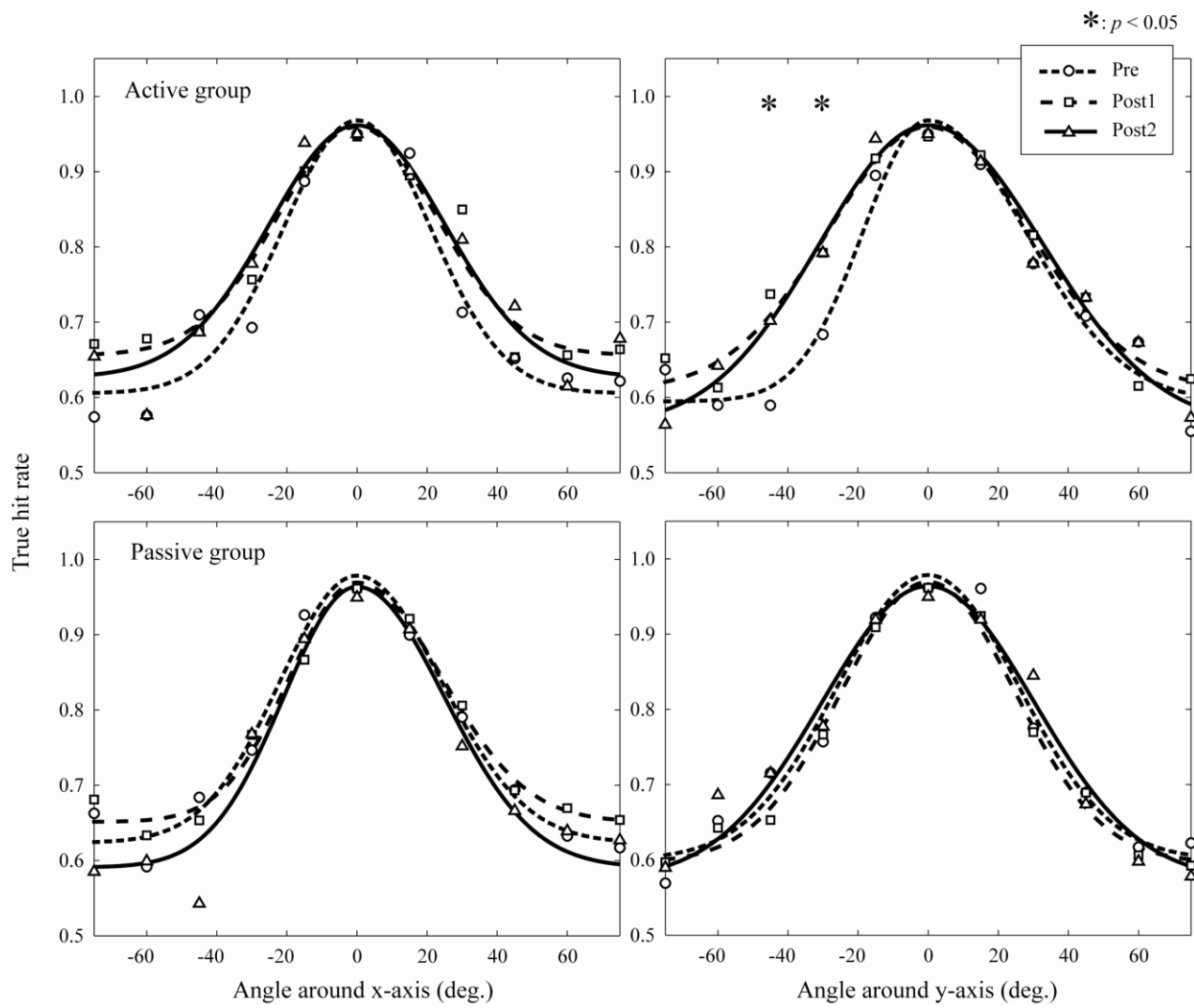


Fig. 10.

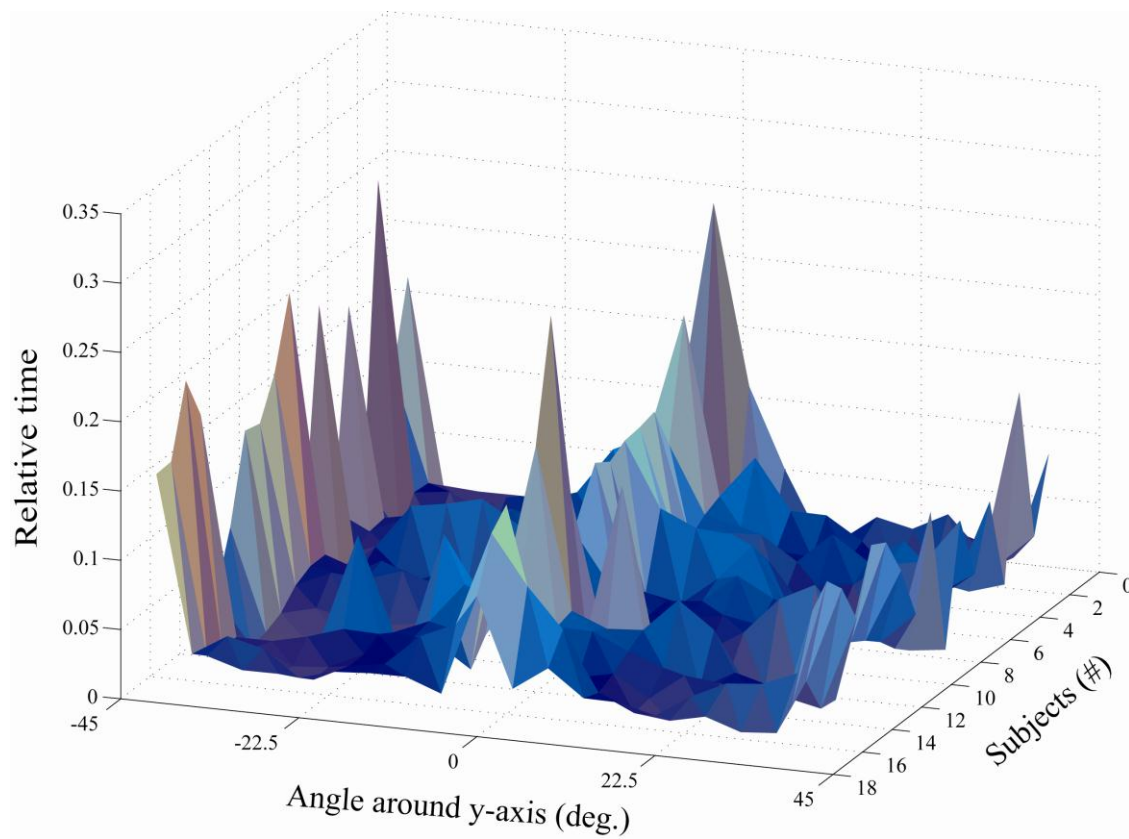


Fig. 11.

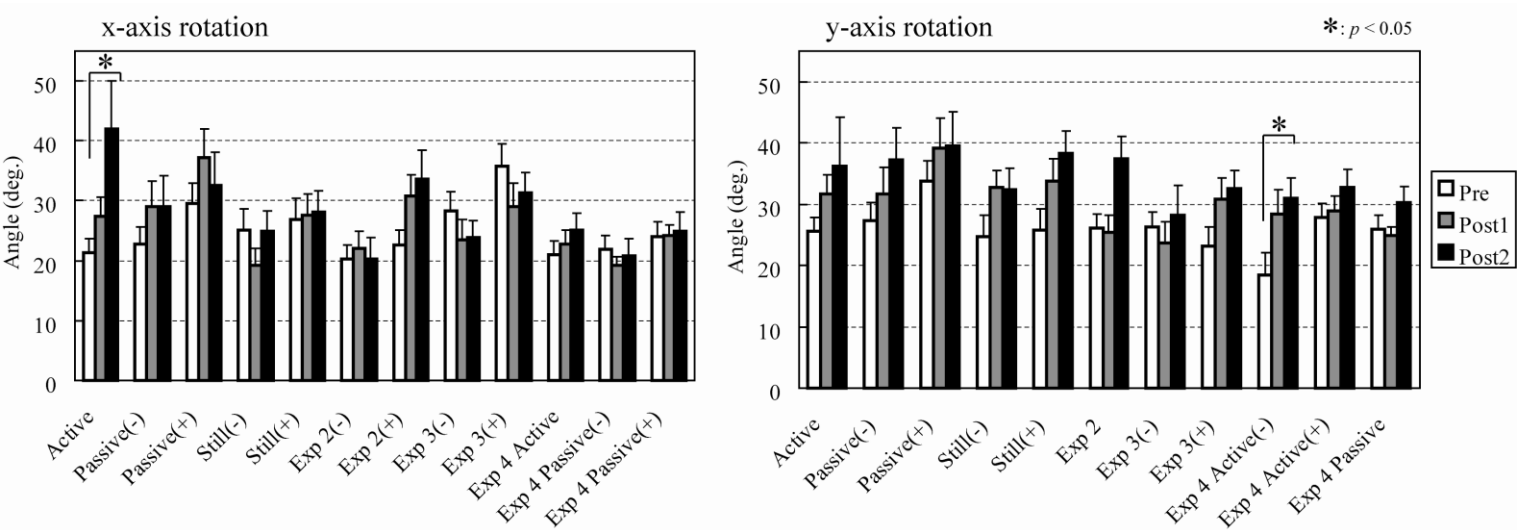


Fig. 12.